Coral classification with hybrid feature representations

CORAL CLASSIFICATION WITH HYBRID FEATURE REPRESENTATIONS

A. Mahmood[⋆]*, M. Bennamoun*[⋆]*, S. An*[⋆]*, F. Sohel*[†]*, F. Boussaid*[⋆]*, R. Hovey*[⋆]*, G. Kendrick*[⋆]*, R. B. Fisher*[∓]

[⋆] The University of Western Australia [†]Murdoch University [∓] University of Edinburgh

ABSTRACT

Coral reefs exhibit significant within-class variations, complex between-class boundaries and inconsistent image clarity. This makes coral classification a challenging task. In this paper, we report the application of generic CNN representations combined with hand-crafted features for coral reef classification to take advantage of the complementary strengths of these representation types. We extract CNN based features from patches centred at labelled pixels at multiple scales. We use texture and color based hand-crafted features extracted from the same patches to complement the CNN features. Our proposed method achieves a classification accuracy that is higher than the state-of-art methods on the MLC benchmark dataset for corals.

Index Terms— corals, deep learning, marine images,classification

1. INTRODUCTION

Coral reefs are vital to marine ecology. Recently, a decline in the health and abundance of coral reefs has been reported [1]. Underwater imaging techniques such as autonomous underwater vehicles (AUV) [2] and towed diver sleds [3] have tremendously increased the amount of coral reef data available for analysis. However, the process of manually annotating this data is cumbersome and inefficient. Automatic underwater image classification is a challenging task because the class boundaries are ambiguous and difficult to define in terms of color, shape or texture. Furthermore, water turbidity and underwater illumination render the images difficult to analyse [4]. As a result, the well-known labelling techniques such as bounding boxes, boundary segmentation and whole image labelling cannot be applied. Instead, marine scientists usually adopt point annotations in practice.

Moreover, a major bottleneck in coral classification is the inherent imbalance of the classes in the datasets due to the abundant presence of non-coral elements such as Crustose Coralline Algae (CCA), turf algae, macroalgae and sand [4]. We have selected the Moorea Labelled Coral (MLC) dataset as it is an excellent benchmarking dataset in coral classification. A sample image from the dataset is shown in Fig. 1. Coral classification can be regarded as a fine-grained classification problem as we want to classify within the sub-categories of corals.

In the recent years, the generic image descriptors extracted from Convolutional Neural Networks (CNN) [5] have replaced the traditional hand crafted features in most of the classification tasks [6]. Image representations extracted from deep CNNs trained on a large dataset such as ImageNet [7] have shown unprecedented successes in diverse classification, localization and recognition tasks [8, 9, 10, 6]. Deep CNNs have a fixed input image size (*i.e.* 224×224). It is important to efficiently resize the input data without information loss when it comes to using CNN based image representations for classification. Multi-scale spatial pooling makes the CNNs more robust

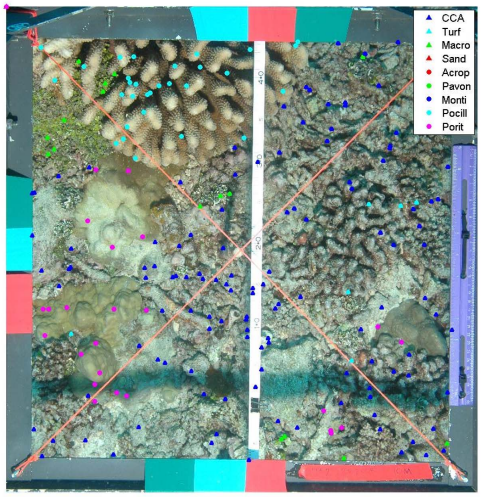

Fig. 1. A sample image from the MLC dataset [4] showing randomly labelled points. The small triangles show non-coral classes, and the small circles represent coral classes.

to different image sizes and has shown to produce a better performance than cropping and warping the input images [11, 12]. **However, due to different annotation techniques in coral images, deep features have not yet been used for the coral reef classification problem.** In practice, marine ecologists prefer random point sampling for ground truth annotations [4]. In this case, random pixels are chosen from the image and are labelled by experts manually. Therefore, the CNN features with spatial pooling cannot be used directly in coral image classification since pixel (instead of bounding box) annotations are usually provided in coral image database.

To make the point-annotated marine data compatible with the input constraints of CNNs, we propose a novel feature extraction scheme based on the Spatial Pyramid Pooling (SPP) [11] approach. Unlike SPP, our approach is local in nature and we call it **local-SPP**. This approach is able to deal with point annotations and class imbalance effectively (Sec. 4).

CNN features learn the image representation given a large amount of training data, while hand crafted features such as SIFT [13] and HoG [14] encode the local patch information about the data. Both of these types of features encode different aspects of the data and hence complement each other. Hand-crafted features are application-dependent and do not require a large number of labelled training data. Furthermore, hand-crafted features commonly encode only one aspect of the data (*i.e.* color, texture, shape) at a time. On the other hand, CNN based features are domain independent, computationally more expensive in training and require a lot of labelled data.

Off-the-shelf CNN features have shown their promise in diverse classification and object detection problems. The main question is whether CNN based features contain sufficient information for coral classification. Deep CNNs are trained on very large datasets such as ImageNet [7] whose images are quite different from the images of the coral datasets in shape and texture. Also, the available coral datasets are not large enough to train a deep CNN for feature extraction from scratch. CNN features learn the representation of data in a supervised fashion independently of the domain area, whereas hand-crafted features encode domain pertinent attributes of data. Moreover, hand-crafted features have shown state-of-art performance for underwater scene classification (Sec. 2). Based on these observations, we propose to combine these two types of features to enhance performance. Our experimental results (Sec. 4) demonstrate that a combination of CNN and hand-crafted features can outperform their individual performances for the coral classification problem.

The main contributions of this paper are: **(1)** the first use of deep features extracted from the VGGNet [15] for coral classification; **(2)** the introduction of a local Spatial Pyramid Pooling (SPP) based technique to improve feature extraction from point annotations; **(3)** the combination of CNN features with texton and color based hand-crafted features for classification improvement.

2. RELATED WORK

For coral classification, researchers have relied on the extraction of color and texture based hand-crafted features for image representation. Marcos *et al*. [16] used Normalized Chromaticity Coordinate (NCC) for color and Local Binary Pattern (LBP) for texture, followed by a 3-layer feed-forward back propagation neural network. However, the NCC features turned out not to be discriminative enough for some coral classes. Stokes and Deane [17] used an RGB histogram and discrete cosine transform based feature vector along with a k-nearest neighbour classifier for benthic coral images. This method is quite fast but the weights of the color and texture features are set manually. Pizaro *et al*. [18] used a feature vector based on NCC histogram, SIFT bag of words and Gabor filter response. Their classification was achieved through voting for the best matches. In their method, each image is classified as one class and the sub-image level classification is not addressed. Beijbom *et al*. [4] introduced the MLC dataset and used a Maximum Response (MR) filter bank followed by texton maps for feature extraction. They also showed that extracting the features in the LAB color space gives a superior performance compared to RGB. They used an SVM classifier with a Radial Basis Function (RBF) kernel for classification.

Image representations extracted from deep CNNs trained on a large dataset such as ImageNet [7] have shown to produce a promising performance for diverse classification and recognition tasks [8, 9, 10, 6]. Spatial pyramid pooling (SPP) [11] and Multi-scale Orderless Pooling (MOP) [12] schemes have made CNNs independent of the input image size and quite robust for diverse classification and recognition applications.

3. PROPOSED METHOD

The proposed method includes four main steps which are demonstrated in Fig. 2. Fig. 2a shows our proposed classification pipeline for CNN based features whereas Fig. 2b shows the classification pipeline for combined features. In this section, these pipelines are described in detail.

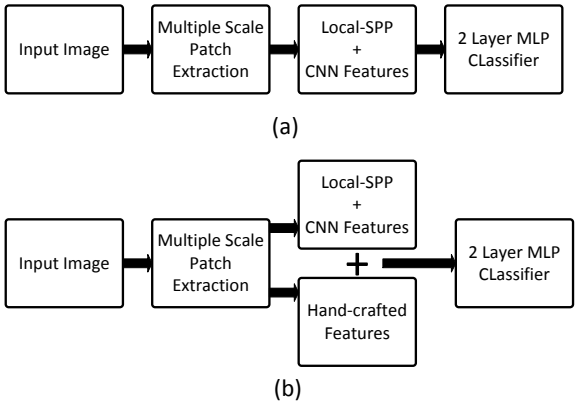

Fig. 2. Block diagram of the proposed method: (a) the pipeline for CNN features; (b) the pipeline for the combined features.

3.1. CNN based feature extraction

Image representations extracted from CNNs trained on large datasets and fine tuned on the domain specific dataset have shown state-of-art performance in scene understanding problems [6]. Marine images in general, and in particular coral reef images, are quite different in nature from the images on which these CNN models are trained. Water turbidity, depth of the imaged location, underwater illumination and color correction are factors which make coral reef data more difficult to classify. Also, marine ecologists prefer annotating the data using pixel labels since the class boundaries are difficult to define.

We have used a pre-trained VGGNet [15] for feature extraction. This network was pre-trained on ImageNet [7] which contains 1000 classes spanned over a million images. VGGnet (configuration D [15]) consists of five convolutional layers and two fully connected layers. Following the approach of [6], we have used the output of the first fully connected layer as the feature vector in our work. The weights of this CNN are then fine tuned using the MLC dataset.

3.2. Local Spatial Pyramid Pooling

VGGnet requries an input image of a fixed size *i.e*. 224×224. As mentioned earlier, the MLC dataset has pixel annotations. We have to manipulate these pixel annotations effectively to meet the input size constraint of the CNN. To do so, we propose to extract square patches of different sizes (multiples of 28) centred at each labelled point (provided with the ground truth) of the MLC dataset. As a result, each image has roughly 200 patches with corresponding ground truth labels. These patches are extracted at different scales to make the feature representation scale invariant. Each patch captures the details in the neighbourhood of a labelled point at a different scale. Each patch is then resized to 224×224 and used as input to the pre-trained VGGnet. The output of the first fully connected layer of this network is used as a feature vector. The final feature vector is obtained by max-pooling these feature vectors as shown in Fig. 3. The resulting feature vector represents the local features in the neighbourhood of the respective pixel. Max-pooling enables us to select the features that express themselves more in the neighbourhood of the corresponding pixel, independent of the scale at which they are extracted.

Note that our patch extraction scheme is spatially pyramidal in nature but it is essentially different from the spatial pyramid pooling (SPP)[11] method. In SPP the whole image is divided into sub-

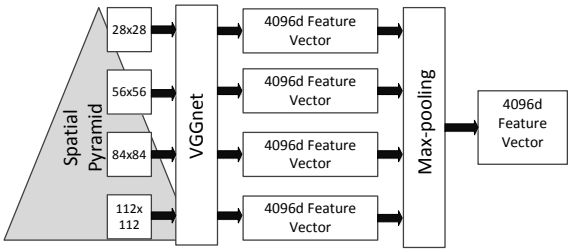

Fig. 3. Local-SPP based feature extraction scheme.

patches and the features over each patch are pooled together to generate one feature representation for the full image. The SPP approach is global in nature and hence the resulting feature vector encodes information of the whole image. However, we are also interested in an efficient representation of the local neighbourhood of the labelled pixel. In our approach, which we have termed "local-SPP", we do not divide the whole image into sub-patches. Instead we extract the patches at different scales centred around the labelled point. The feature vectors acquired from multi-scale patches are then max-pooled together to find the maximum scale-invariant response. Therefore, the resulting feature vector encodes the most prominent features in the proximity of the corresponding pixel.

3.3. Combined features (CF)

Hand-crafted features are quite popular when it comes to marine data analysis (Sec. 2). However, in the past few years, CNN features have shown to outperform state-of-art hand crafted features for many computer vision problems. CNN features are extracted from deep networks that are trained on a large number of RGB images. Beijbom *et al.* [4] showed that RGB color space is not effective enough for the encoding of the color information of underwater imagery. To alleviate this bottleneck, we propose to combine CNN features with hand-crafted features to increase the classification performance. We have used the color and texton based features of [4] for all of our experiments (Sec. 4). The 4096-dimensional feature vector is concatenated with the 540-dimensional feature vector of [4] to obtain a 4636-dimensional combined feature vector.

3.4. Classifier

VGGnet is a very large network with approximately 140 million parameters [15]. Hence, it is not feasible to train it from random initializations since our dataset is quite small compared to ImageNet. We therefore rely on pre-trained features [6]. For classification, we use a two layer Multilayer Perceptron (MLP) network trained using the MLC dataset. The architecture of this network is shown in Fig. 4. The MLP network consists of two fully connected layers followed by a soft-max layer with 9 output classes. This network architecture is similar to the fully connected portion of the VGGnet except that the last layer has only 9 classes. The parameters a and b were optimized using the training data of experiment 1 (Sec. 4) since it has the smallest training set. We used the same parameters for the other two experiments.

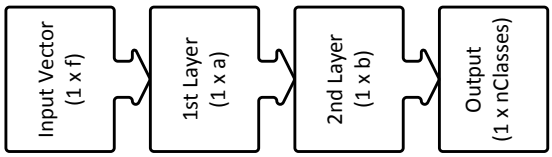

Fig. 4. The MLP architecture used for classification.

4. EXPERIMENTS AND RESULTS

4.1. Dataset

The MLC dataset [4] contains 2055 images collected over three years: 2008, 2009 and 2010. It also contains random point annotation (x, y, label) for the nine most abundant labels, four non coral labels: (1) Crustose Coralline Algae (CCA), (2) Turf algae, (3) Macroalgae and (4) Sand, and five coral genera: (5) Acropora, (6) Pavona, (7) Montipora, (8) Pocillopora, and (9) Porites. MLC dataset contains 400,000 expert pixel annotations and an inherent class imbalance, which makes it a challenging dataset.

4.2. Feature extraction

We use the output of the first fully connected layer of VGGnet as a feature vector in all of our experiments. The input to the network is resized to 224×224 pixels which gives an output feature vector of 4096 dimensions. In our proposed method, we crop the patches of sizes that are multiples of 28 centred at the labelled points (given with the ground truth data). These patches are then resized to 224 x 224 pixels before using them as input to the VGGnet. We extracted patches of 4 different sizes for our experiments i.e., 28×28, 56×56, 84×84 and 112×112. Since these patches are centred around the point labels, the feature vector encodes the neighbourhood information at different scales. We proceed from a coarse to fine representation as we reduce the patch size. When features at different levels are max-pooled together, the feature vector encodes the maximum responses in the neighbourhood of a pixel and retains the best information independently of scale. This accounts for the increase in the performance of the classifier for local-SPP features.

We evaluated the classification performance of our algorithm by performing the same three experiments reported in [4]. **In experiment 1**, the classifier is trained on two-thirds of the images from the year 2008 and tested on the remaining images from the same year. **In experiment 2**, the images from year 2008 are used for training and the images from 2009 constitute the test set. **In experiment 3**, the training set consists of images from the years 2008 and 2009, whereas the test set consists of images from year 2010. The coral reef was imaged using the same protocol over the three years.

4.3. Hand crafted features

Beijbom *et al.* [4] introduced an algorithm based on color and texture descriptors over multiple scales. They used a Maximum Response (MR) filter bank along with the color information in the LAB color space, followed by the extraction of texton maps at multiple scales. Their algorithm performed better than other texture based descriptors for coral classification. We have used their results as a benchmark for our algorithm. The texton based feature vector extracted in this way are not linearly separable and hence they have used an SVM classifier with a Radial Basis Function (RBF) kernel for classification. To overcome the class imbalance problem, the

Features	Exp 1	Exp 2	Exp 3
Hand-crafted features in RGB [4]	72.7	66	80.8
Hand-crafted features in LAB [4]	74.3	67.3	83.1
Local SPP + CNN features	77.4	69.2	82.8
Combined Features (CF)	**77.9**	**70.1**	**84.5**
# of Training Samples	87,428	131,260	263,372
# of Test Samples	43,832	132,112	129,927

Table 1. Overall classification accuracies for different feature representations

training data was down-sampled and the cost function was assigned a weight that is inversely proportional to the down-sampling rate.

4.4. Classification Results

Beijbom *et al*. [4] showed that pre-processing the coral images in LAB color space gives a better performance than RGB, HSV and gray color spaces. VGGnet is trained using a large number of RGB images and hence cannot be used with LAB color-space. To overcome this limitation, we combined the hand-crafted features of Beijbom *et al*. with CNN features which results in a hybrid feature representation. The 4096 dimensional CNN feature vector is concatenated with the 540 dimensional texton based feature vector to obtain a combined feature vector. The individual feature vectors are normalized prior to concatenation. The overall classification accuracies are shown in Table 1. The accuracies for the hand-crafted features in [4] are reported in the first two rows for RGB and LAB color-spaces respectively. For the first two experiments, the local-SPP based CNN features achieve a better performance compared with the respective results of the first two rows. For the third experiment, the local-SPP based CNN features produce a better performance only compared to the RGB ones and perform slightly lower compared to the LAB ones. It can be seen that the combined features approach outperforms all the other approaches. The CF approach achieved an accuracy of 77.9%, 70.1% and 84.5% in the three experiments (compared to the corresponding 74.3%, 67.3% and 83.1% classification accuracy in [4]). Furthermore, the combined features perform significantly better than the local-SPP based CNN features when the classifier is trained with a large amount of data (*i.e*. experiment 3). The last two rows of Table 1 give the number of training and test samples used in the corresponding experiment.

We also calculated the Average Class Precision (ACP) for our experiments. ACP is the mean of the diagonal of a confusion matrix. Higher ACP implies a better classification. Table 2 shows a comparison of ACP of our experiments with those of [4]. Our local-SPP features and combined features have both a higher ACP than those reported in [4]. The combined features performed the best. Our improved classification accuracy and ACP demonstrates that the local-SPP-based features and combined features addressed the problem of class imbalance more effectively. In the case of corals, the most abundant class overshadows the lesser classes when the patches are extracted at one scale. Since we extract patches at different scales and then max-pool them, this makes the less abundant classes more prominent in the resulting feature vectors. This helps the classifier to cope with the inherent class imbalance problem effectively.

4.5. Sub-classification within classes

There are three important sub-classification tasks within MLC dataset: **(1)** the binary classification task between corals and non-corals, **(2)** within coral classification *i.e*. if a pixel is classified as coral, we need to find its coral type and **(3)** within non-corals *i.e*. if a pixel is classified as a non-coral, we need to sub-classify it within the non-coral classes. Table 3 shows the detailed results for these sub-classification tasks using combined features and compares them with the results of [4]. Our results show that CF approach performs better than the baseline results of [4] for all experiments. The improved diagonals of the confusion matrices of our method foreshadowed the results of Table 3. Since CCA is present in abundance, the last sub-classification task (*i.e*. within non-corals) is quite challenging in itself. This fact accounts for the comparatively low classification accuracies in the last column of Table 3.

Features	Exp 1	Exp 2	Exp 3
Hand-crafted features in LAB [4]	0.60	0.56	0.60
Local SPP + CNN features	0.67	0.62	0.66
Combined Features (CF)	**0.69**	**0.63**	**0.68**

Table 2. Average class precision for different feature representations

	Corals and non-corals	Within corals	Within non-corals
Exp 1 [4]	92	97	78
Exp 1 [CF]	**94**	**98**	**79**
Exp 2 [4]	93	91	70
Exp 2 [CF]	**95**	**92**	**72**
Exp 3 [4]	95	97	87
Exp 3 [CF]	**96**	**98**	**88**

Table 3. Results for sub-classification tasks

5. CONCLUSION

In this work, we reported the first ever application of deep learning to the coral reef classification problem. We proposed to use pre-trained CNN representations extracted from VGGnet with a 2 layer MLP classifier (trained with the MLC dataset) to address coral classification. We investigated the effectiveness of transferring the feature representations learned from deep CNNs trained on ImageNet to coral reef data. We also introduced a local-SPP approach to extract features at multiple scales to deal with the ambiguous class boundaries of corals. We then combined CNN based features with texton and color based hand-crafted features for a better classification performance. Our experiments confirmed that our proposed method achieved a state-of-art classification accuracy on the MLC dataset. Furthermore, our method deals with the class imbalance problem effectively.

6. REFERENCES

[1] Forest Rohwer, Merry Youle, and Derek Vosten, *Coral reefs in the microbial seas*, vol. 1, Plaid Press United States, 2010.

[2] MR Patterson and NJ Relles, "Autonomous underwater vehicles resurvey bonaire: a new tool for coral reef management," in *Proceedings of the 11th International Coral Reef Symposium*, 2008, pp. 539–543.

[3] Jean C Kenyon, Russell E Brainard, Ronald K Hoeke, Frank A Parrish, and Casey B Wilkinson, "Towed-diver surveys, a method for mesoscale spatial assessment of benthic reef habi-

tat: a case study at midway atoll in the hawaiian archipelago," *Coastal Management*, vol. 34, no. 3, pp. 339–349, 2006.

[4] Oscar Beijbom, Peter J Edmunds, David Kline, B Greg Mitchell, David Kriegman, et al., "Automated annotation of coral reef survey images," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1170–1177.

[5] B Boser Le Cun, John S Denker, D Henderson, Richard E Howard, W Hubbard, and Lawrence D Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*. Citeseer, 1990.

[6] Ali S Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson, "Cnn features off-the-shelf: an astounding baseline for recognition," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*. IEEE, 2014, pp. 512–519.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 248–255.

[8] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," *arXiv preprint arXiv:1310.1531*, 2013.

[9] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jagannath Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 580–587.

[10] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1717–1724.

[11] Yunchao Gong, Liwei Wang, Ruiqi Guo, and Svetlana Lazebnik, "Multi-scale orderless pooling of deep convolutional activation features," in *Computer Vision–ECCV 2014*, pp. 392–407. Springer, 2014.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Computer Vision–ECCV 2014*, pp. 346–361. Springer, 2014.

[13] David G Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Ieee, 1999, vol. 2, pp. 1150–1157.

[14] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.

[15] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[16] Ma Shiela Angeli Marcos, Maricor Soriano, and Caesar Saloma, "Classification of coral reef images from underwater video using neural networks," *Optics express*, vol. 13, no. 22, pp. 8766–8771, 2005.

[17] M Dale Stokes and Grant B Deane, "Automated processing of coral reef benthic images," *Limnology and Oceanography: Methods*, vol. 7, no. 2, pp. 157–168, 2009.

[18] Oscar Pizarro, Paul Rigby, Matthew Johnson-Roberson, Stefan B Williams, and Jamie Colquhoun, "Towards image-based marine habitat classification," in *OCEANS 2008*. IEEE, 2008, pp. 1–7.